

Distribution Models and a Dated Phylogeny for Chilean *Oxalis* Species Reveal Occupation of New Habitats by Different Lineages, not Rapid Adaptive Radiation

CHRISTOPH HEIBL AND SUSANNE S. RENNER*

Systematic Botany and Mycology, University of Munich, Menzinger Straße 67, 80638 Munich, Germany;

* Correspondence to be sent to: Institute of Systematic Botany and Mycology, Menzinger Straße 67, 80638 Munich, Germany;
E-mail: renner@lrz.uni-muenchen.de.

Abstract.—Among the World's most challenging environments for plant life is the Atacama Desert, an arid zone extending over 1300 km and from sea level to 2000/3000 m altitude along the southwestern Andean foothills. Plants there and in the adjacent Mediterranean zone exhibit striking adaptations, and we here address the question whether in a species-rich clade such adaptations arose in parallel, at different times, or simultaneously. Answering this type of question has been a major concern of evolutionary biology over the past few years, with a growing consensus that lineages tend to be conservative in their vegetative traits and niche requirements. Combined nuclear and chloroplast DNA sequences for 112 species of Oxalidales (4900 aligned nucleotides) were used for a fossil-calibrated phylogeny that includes 43 of the 54 species of Chilean *Oxalis*, and species distribution models (SDMs) incorporating precipitation, temperature, and fog, and the phylogeny were used to reconstruct ancestral habitat preferences, relying on likelihood and Bayesian techniques. Since uneven collecting can reduce the power of SDMs, we compared 3 strategies to correct for collecting effort. Unexpectedly, the *Oxalis* flora of Chile consists of 7 distant lineages that originated at different times prior to the last Andean uplift pulse; some had features preadapting them to seasonally arid or xeric conditions. Models that incorporated fog and a "collecting activity surface" performed best and identified the Mediterranean zone as a hotspot of *Oxalis* species as well as lineage diversity because it harbors a mix of ancient and young groups, including insufficiently arid-adapted species. There is no evidence of rapid adaptive radiation. [Atacama Desert; divergence time estimation; Mediterranean climate; species distribution modeling; species distribution models; Southern Andes; target-group background sampling.]

Among the World's most challenging environments for plant life is the Atacama Desert. This arid zone extends from the border between Chile and Peru at 18°S over 1300 km to about 30°S. It encompasses a stretch of land from sea level to 2000 or 3000 m altitude (depending on latitude) along the western Andean foothills (Luebert and Pliscoff 2006). Toward the South, the desert transitions into Mediterranean shrub land and then steppe scrub. The Atacama itself includes absolute desert and desert scrub, depending on the influence of the rain shadow climate that dominates central Chile. The plants that occur in the Atacama Desert possess striking adaptations to the arid conditions, such as stunted growth, minute leaves, deep roots, or water storage organs.

Did such adaptations evolve in situ (as the conditions changed) or were they rather preadaptations in lineages living elsewhere and that then moved in? Answering this type of question has been a major concern of evolutionary biology over the past few years, with a growing consensus among botanists being that lineages tend to be conservative in their vegetative traits and niche requirements. As Donoghue (2009: 11549) has put it, "it may often be easier for (plant) lineages to move than it is for them to evolve." There is ample evidence supporting Donoghue's view (Peterson et al. 1999; Webb 2000; Prinzing et al. 2001; Ackerly 2004; Wiens and Graham 2005; Crisp et al. 2009). Such niche conservatism implies that the migration of preadapted species into a newly formed habitat from surrounding areas must have played a large role in the assembly of communities.

Most plant groups that occur in the Atacama Desert are represented there by one or a few species, and only

a handful of genera are diverse, including *Heliotropium* section *Cochranea* (17 species; *Heliotropiaceae*; Luebert and Wen 2008), *Nolana* (71 coastal endemics from Peru to Chile; *Solanaceae*; Dillon et al. 2009), and *Oxalis*, the focus of this study. *Oxalis* (*Oxalidaceae*), the woodsorrel genus, comprises about 480 species worldwide, including many in Europe, North America, Africa, and the Neotropics. In Chile, *Oxalis* comprises 54 species and is among the most diverse flowering plant genera (Fuentes et al. 1995; our Fig. 1). Its sectional classification is based mainly on growth forms (Lourteig 2000) and implies that the species in Mediterranean Chile do not form a clade. This sets up the hypothesis that *Oxalis* lineages in Chile adapted to more or less arid climates in parallel. Their diverse and obvious morphological adaptations, especially in growth habit (Fig. 1), and the relatively well-constrained origin of different climatic conditions in Chile make this a promising system in which to study the coincidence between the availability of a particular habitats and its occupation.

Although the Atacama Desert may date back to 150 Ma (Hartley et al. 2005; Clarke 2006), its hyperaridity dates to the mid-Miocene (15–10 Ma), when the raising Andes began to block moist air from the Amazon basin (Alpers and Brimhall 1988; Houston and Hartley 2003; Rech et al. 2010). Previously, year-round precipitation had been provided by easterlies from the Amazon during the summer and westerlies from the ocean during the winter. Further south, the uplift of the Patagonian Andes caused the development of the Patagonian steppes and the Monte desert in Argentina (Blisniuk et al. 2005). These arid habitats west and east of the Andes,

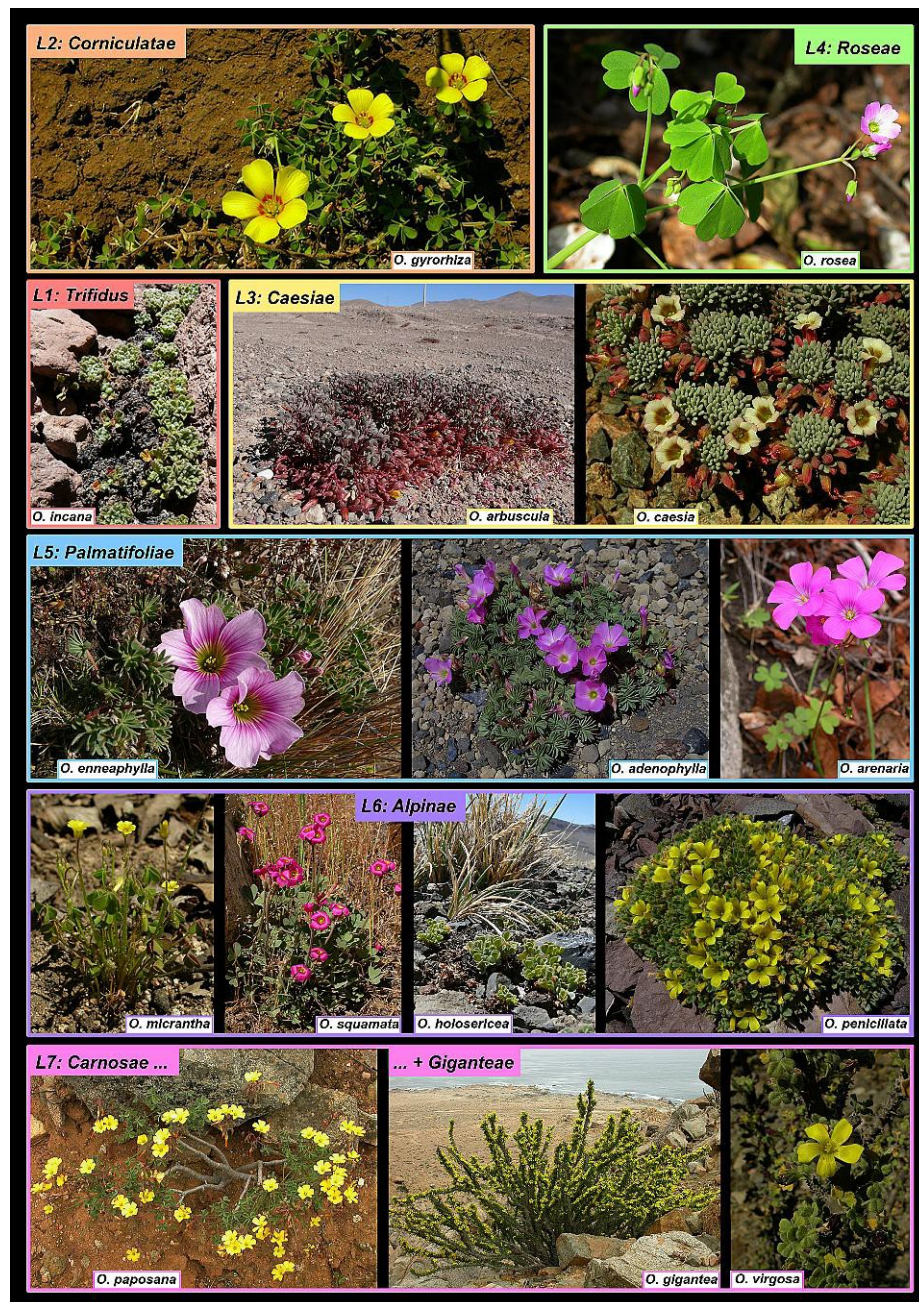


FIGURE 1. Growth forms found in the Chilean *Oxalis* species analyzed in this study. The 7 main *Oxalis* lineages are numbered and color coded as in the phylogeny (Fig. 2) and in Table 1, which lists their divergence time estimates. Photos by C. Heibl.

together with newly created (also water-stressed) alpine environments above timberline, provided ample opportunity for the evolution and diversification of drought-adapted lineages.

Dependence of species on particular climatic conditions can be inferred from species distribution models (SDMs) derived from georeferenced collections. A problem for such modeling is the bias introduced by uneven collecting (Phillips and Dudík 2008; Phillips et al. 2009). If this bias is ignored, as under the default setting in widely used spatial modeling programs, SDMs

will reflect collecting activity rather than true species distributions. The bias can be reduced by background (pseudo-absence) data not sampled at random but instead spatially biased toward easily accessed areas just like the presence data. For this, one needs a distribution of collecting intensity relevant to the target group, which can be obtained from herbarium collections of the larger taxon to which the focal species belongs. Collecting in the Atacama Desert and the Andes is concentrated along roads, whereas in southern Chile, it is highly concentrated along the seashore. Comparing the effects

of different background sampling corrections therefore seemed particularly important.

This study thus uses sophisticated SDMs for the Chilean species of *Oxalis*, together with a dated phylogeny of the genus, to answer the following questions: (i) How does the timing of *Oxalis* diversification in southern South America relate to the Andean orogeny? (ii) Are switches between habitat types occupied by *Oxalis* (forest, shrub, desert, steppe, and alpine habitats) congruent with the temporal origin of the respective habitats? and (iii) Is Chile's Mediterranean zone a refuge (sink) or a source of new *Oxalis* lineages adapting to new (post-Miocene) habitats?

MATERIALS AND METHODS

Taxon Sampling

We sampled 43 of the 54 species of *Oxalis* species known from Chile (and adjacent countries). *Oxalis perdicaria* is widespread on both sides of the Andes in seasonal dry habitats, and *O. magellanica* occurs in high-altitude and high-latitude bogs in Patagonia, Tasmania, and New Zealand. Species names and authors, herbarium vouchers, their deposition, and the geographic origin of all included accessions are shown in online supplementary material Table S1. Sequences from GenBank were added for species from *Oxalis* sections *lonoxalis*, *Pseudobulbosae*, *Corniculatae*, the north-central Andean *O. tuberosa* alliance, and species from the Greater Cape Floristic Region (GCFR). Plant identifications were made or checked by the 1st author during work in the relevant herbaria.

Representatives of Huaceae, Cephalotaceae, Brunelliaceae, Elaeocarpaceae, Cunoniaceae, and Connaraceae as well as one species each of *Averrhoa*, *Biophytum*, *Daphnia*, and *Sarcotheca* (*Oxalidaceae*) were used as outgroups. This broad outgroup sampling was needed for constraining the molecular clock model with fossils (see below).

DNA Isolation and Sequencing

Total genomic DNA was isolated from silica-dried leaf samples with a modified hexadecyltrimethylammonium bromide procedure (Bahnweg et al. 1998). Of plasmid DNA, we sequenced the *rbcl* gene (using standard primers), the *trnL* intron and *trnL*–*trnF* intergenic spacer (Taberlet et al. 1991), the *psbA*–*trnH* intergenic spacer (Sang et al. 1997), and the *trnS*–*trnG* intergenic spacer (Hamilton 1999). Of nuclear DNA, we sequenced the internal transcribed spacers (ITS) regions and the intervening 5.8S rRNA gene using standard primers. For the large *Oxalidales* taxon set used for molecular clock dating, we only used 3 of these markers (*rbcl*, *trnL*–*L-F*, and ITS; 25% missing nucleotides). Polymerase chain reaction conditions, product cleaning, and cycle sequencing followed standard protocols. Forward and reverse sequences were edited with Sequencher 4.6 (Gene Codes, Ann Arbor, MI).

Alignments and Phylogenetic Analyses

DNA sequences were aligned using PRANK (Löytynoja and Goldman 2005). This alignment program avoids penalizing single insertion events multiple times by keeping track of sequence sites that were within a gap in previous pair wise sequence comparisons while moving along its guide tree. Therefore, it is well suited for noncoding sequences, which make up the majority of our data. Divergent and ambiguously aligned regions were removed with GBLOCKS (Talavera and Castresana 2007) using medium-stringency conditions, i.e., only blocks of a minimum length of 8 nucleotides with a minimum of 50% of sequences for a conserved position and 85% of sequences for a flank position were retained in the alignments; the maximum number of contiguous nonconserved nucleotide positions was 3; gaps were allowed.

Phylogenetic analyses relied on maximum-likelihood (ML) as implemented in RAxML 7.0.4 (Stamatakis 2006) and used the GTR + Γ model of sequence evolution. Modeltest (<http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html>) selected this model as the best fitting for most data partition except *rbcl* (K81 + Γ) and *trnS*–*trnG* (GTR). We also ran partitioned analyses that allowed the combined chloroplast data and the nuclear ITS to have distinct substitution rates. This did not result in topological differences or higher node support. Support for the ML topology was assessed with the rapid bootstrap algorithm implemented in RAxML using 100 replicates (Stamatakis et al. 2008).

Molecular Clock Dating

As there are no reliably identified old *Oxalis* fossils, we calibrated genetic distances by using fossils from a related clade and by setting the split between *Oxalis* and *Averrhoa* to 47 ± 4 Ma based on an estimate by Wikström et al. (2001). Wikström et al. (2001) employed nonparametric rate smoothing on a data set of chloroplast and nuclear genes (4733 aligned bp) for 560 angiosperms and 7 outgroups, using the 84 Ma split between Fagales and Cucurbitales as the calibration point. The fossil calibrations are from the Cunoniaceae. Late Santonian–Early Campanian flowers assigned to *Platydiscus*, an extinct Cunoniaceae (Schönenberger et al. 2001), provide a minimal age for the split of Cunoniaceae from its sister clade. *Platydiscus* is from the Åsen formation in Sweden, from a horizon at the Santonian/Campanian boundary (magnetostratigraphical dating) that is about 83.5 ± 0.7 Ma old (Friis E.-M., personal communication). The other fossils are leaves of *Eucryphia* (Cunoniaceae) from Tasmania, dated to 55.8 Ma (Barnes and Jordan 2000). Since the cuticle morphology of *Eucryphia* is characteristic of the genus (Jordan G., personal communication), these leaves provide a minimal constraint for the split of *Eucryphia* from the remaining Cunoniaceae.

We used the uncorrelated-rates approach implemented in BEAST 1.5.2 (Drummond et al. 2006; Drummond and Rambaut 2007), which uses Bayesian Markov

chain Monte Carlo (MCMC) runs to coestimate topology and node ages. Posterior probability distributions of node ages were obtained from the 112-taxa-6-marker alignment. We used a pure birth process tree prior and the ML tree as a starting tree, after making it ultrametric by using penalized likelihood (Sanderson 2002) as implemented in Analysis of Phylogenetics and Evolution with a lower bound of 55.8 Ma on the split between *Eucryphia* and *Acrophyllum*.

BEAST runs used a normal distribution (mean = 47 Ma, SD = 4) for the secondary age calibration from Wikström et al. (2001) and lognormal distributions for the 2 fossil-calibrated nodes. The latter's zero offset was set equal to the age estimate of the corresponding fossils, and the 95% confidence interval (CI) was chosen so as to include 10 myr before the fossil age. Broader CIs that included 20 myr before the fossil age did not significantly change time estimates. Eight MCMC chains were run for 8–10 million generations, with parameters sampled every 10,000th step. Results from individual MCMC chains were combined and convergence checked using TRACER 1.4.1 (Rambaut and Drummond 2007). Effective sample sizes for all relevant estimated parameters and node ages were well above 200. We also ran an analysis without the data in order to assess their informativeness. This analysis indicated that the effective priors were similar to the original priors and that the posteriors obtained after adding the data departed from both.

Species Distribution Modeling

SDMs were constructed with MAXENT 3.3.3 (Phillips et al. 2006), which generates random pseudo-absences and thus requires only presence data, making it a suitable tool for analyzing collection-based data. MAXENT uses a maximum entropy approach to predict landscape suitability for a species as a function of environmental variables, usually the 19 bioclimatic variables and the digital elevation model in the WorldClim data set (Hijmans et al. 2005). These variables represent the means and variation in temperature and precipitation. An important feature not captured is the amount and frequency of coastal fog, which is highly correlated with the occurrence of “fog oases” (Rundel et al. 1991), where the coastal mountain range intercepts marine stratocumulus clouds (Cereceda et al. 2002). In the absence of comprehensive measurements of the amount and frequency of fog in northern Chile (Cereceda and Schemenauer 1991; Larrain et al. 2002), we modeled the distribution of fog as a categorical variable based on topographical features (i.e., altitude, slope, aspect, ocean proximity) derived from the GTOPO30 digital elevation model (USGS 1996). The combination of these topographical features occurs within 10 km of the coast and between 400 and 1200 m altitude, with a slope of 8° or higher and with a predominantly western to southern exposition.

Species occurrence data came from the 1st author's fieldtrips to Chile and Peru in October/November 2007 and January/March 2008 and the herbarium collec-

tions housed at the Museo Nacional de Historia Natural in Santiago (SGO), the Universidad de Concepción (CONC), the Royal Botanic Gardens (K), the Royal Botanic Garden Edinburgh (E), and literature. We used the Biogeomancer workbench (<http://bg.berkeley.edu/latest>) to georeference all records with locality descriptions sufficiently detailed to obtain a radius of geographical uncertainty of less than 5 km. This yielded 2016 georeferenced data points, 1779 of which could be assigned to native species. For 4 species with fewer than 9 collections and high spatial autocorrelation of sampling points, we approximated distribution models in GRASS GIS as the intersection between a buffer of 2–10 km around sampling points and a decisive environmental condition, namely the occurrence of coastal fog (*O. leucophylla*, *O. novemfoliolata*, and *O. morenoensis*) or an altitude of >1000 m (*O. campanensis*). Forty-two more widespread species were represented by 1764 localities, ranging from 9 to 147 localities per species (median = 33); these were modeled with MAXENT.

We compared 4 strategies to deal with sampling bias: (i) Using an overlapping target-group background (Phillips et al. 2009), i.e., the entire set of 2016 localities of *Oxalis* (including introduced and undetermined specimens); (ii) converting the coordinates of the localities into a raster layer representing the number of accessions per grid cell (NAPG) and then using this as a bias file in MAXENT (a bias file determines the sampling distribution according to which 10⁴ background points are drawn); (iii) deriving a continuous collecting activity surface (CAS) by natural neighbor (nn) interpolation from the NAPG layer with the nn library and the GRASS GIS add-on *r.surf.nn*bathy (Sieczka 2008; Sakov 2010) and then using this as a bias file in MAXENT; and (iv) randomly sampling 10⁴ background points according to a uniform distribution (UNI, the default in MAXENT).

Binary (presence–absence) distribution maps for each species were obtained by combining landscape units with MAXENT probabilities greater than the maximum training sensitivity plus specificity (Jiménez-Valverde and Lobo 2007). Summation of the individual binary distribution maps provided an estimate of *Oxalis* species numbers per landscape unit. Analyses (except for the maximum entropy modeling) were done in the R language (R Development Core Team 2008) and GRASS GIS (GRASS Development Team 2009).

Inference of Ancestral Habitats

To infer ancestral habitat preferences, we used ML and Bayesian trait reconstruction. This required binning habitat types into a limited number of categories, here forest, shrub, desert, alpine, and steppe. The category “forest” (mostly present in outgroups) included sclerophyllous forest under Mediterranean or monsoon climates but also temperate forest and bogs. The category “shrub” comprised Mediterranean-type shrub land, such as Chilean matorral and the South African fynbos. Precipitation in these shrublands is intermediate

between sclerophyllous forest and desert. The category “desert” included habitats with severe drought stress (rainfall unpredictable and <300 mm/year) found in the South American Atacama Desert and the South African Karoo-Namib Desert. The “alpine” habitat included plant communities above timberline, such as *Festuca* steppes and cushion or rosette plant communities. The “steppe” category included the cold temperate Patagonian steppe, which is characterized by harsh winters and aridity due to the rain shadow of the southern Andes, intensified by high wind velocities.

Ancestral state reconstruction relied on the ultrametric phylogenies obtained with BEAST. We thinned the posterior sample to 1000 trees and extracted the *Oxalis* clade from each tree. To account for the weak support of basal nodes, we iterated trait change inference over the 1000 trees. The program MULTISTATE in the BAYESTRAITS package (Pagel 1994) was used to fit different continuous-time Markov models of trait change to the data. Because of exponential increase in rate parameters with increasing number of states, we compared a four-state model with a five-state model by merging the categories shrub and forest. Following Pagel’s rule of thumb that for each rate parameter to be estimated, there should be 10 species with known character states, we inferred between 1 and 5 rate parameters for the five-state model and between 1 and 6 parameters for the four-state model. The best-fit model was chosen based on the corrected Akaike Information Criterion (AIC_c).

For Bayesian ancestral state reconstruction, we used a reversible jump Markov chain Monte Carlo method (Pagel and Meade 2006), contained in the BAYESTRAITS package, which automatically finds the posterior distribution of models if provided with the appropriate choice of priors (Pagel et al. 2004). This yields posterior distributions of rate coefficients and state probabilities, incorporating phylogenetic uncertainty as well as uncertainty in the ancestral state reconstruction. Following the manual, we explored different combinations of the *ratedev* parameter, i.e., the deviation of the normal distribution from which rates are drawn and the hyperpriors/priors to obtain acceptance rates of roughly 20%. For the final runs, we set the *ratedev* parameter to 8 and used exponential priors and gamma priors with their parameters seeded from uniform hyperprior distributions between 0 and 20.

RESULTS

Phylogenetic Analyses and Divergence Times

The inferred divergence times suggest an Eocene origin for the *Oxalis* crown group (Fig. 2, Table 1, which also list the confidence intervals). The divergence times of the 7 lineages fall between the Middle Eocene (Table 1: L3 section *Caesia*) and the Late Miocene. The *O. rosea* clade (L4) split from the North American/Eurasian *O. acetosella* clade in the Middle Miocene. Arid Chile harbors 7 *Oxalis* lineages (highlighted in Fig. 3) henceforth referred to as L1 through L7. Species richness is

concentrated in lineages L5–L7 and appears to relate to geographic distribution and morphology (Table 2 and Discussion Section). There is a good overall agreement between the molecular phylogeny and the most recent sectional classification (Lourteig 2000), suggesting that the 11 Chilean *Oxalis* species that we did not sequence may also group according to their sectional placement.

Inferred Ancestral Habitats

With five habitat preference states (Materials and Methods section), the equal rates model had the best average AIC_c values ($Y_1 = -153.60$, $SD_1 = 4.75$) over the 1000 tree replicates compared with the two-rate and five-rate model ($Y_2 = -149.69$, $SD_2 = 5.10$; $Y_5 = -134.71$, $SD_5 = 4.40$; $F = 4390$, $P < 0.001$). With four habitat preference states results were similar. Adding more than one rate parameter to the model did not improve model fit ($F = 8215$, $P < 0.001$). In the Bayesian trait reconstruction, the reversible jump MCMC chain most often visited models with 2 rate parameters, followed by models with one rate parameter in the case of the four-state model and 3 rate parameters for the five-state model. Even with these comparatively sparsely parameterized models, the algorithm failed to reconstruct ancestral states for deeper nodes with probabilities departing significantly from random (data not shown).

Reconstruction of ancestral habitats on the maximum clade credibility tree from the 6-marker data set (Fig. 4) suggests that ancestral *Oxalis* occurred in forest and suggests an evolutionary trend toward the occupation of drier (shrub) habitats in the “arid clade” (so labeled in Fig. 4). Several shifts from forest to shrub habitats (color coded in Fig. 4) predate the last Andean uplift pulse (2nd pale blue band in Fig. 4) and thus the establishment of the Mediterranean climate. Only *O. megalorrhiza* and *O. tortuosa* shifted to dry shrub habitat after the establishment of the Mediterranean climate. The mid-Miocene onset of hyperaridification of the Atacama (yellow bands in Fig. 4) correlates with the diversification of 2 small clades (one with 15 species and the other with 2), whereas the alpine-adapted species (pale blue squares in Fig. 4) all post-date the final Andean uplift pulse (broad blue band in Fig. 4). The only steppe-adapted species (the 2 lilac squares Fig. 4) date to the Pliocene, when this habitat would already have been available for at least 8–10 myr.

Performance of the Different SDMs

Of the four strategies to deal with sampling bias, the first, using an overlapping target-group background, performed worst, probably because we had 2016 localities rather than the recommended 10^4 (Phillips et al. 2009). The 2 strategies, NAPG and CAS, performed nearly equally well, except that CAS predicted less into uninhabitable habitats, such as the core Atacama Desert and Patagonian ice shields. The 4th strategy, randomly sampling background points according to a uniform

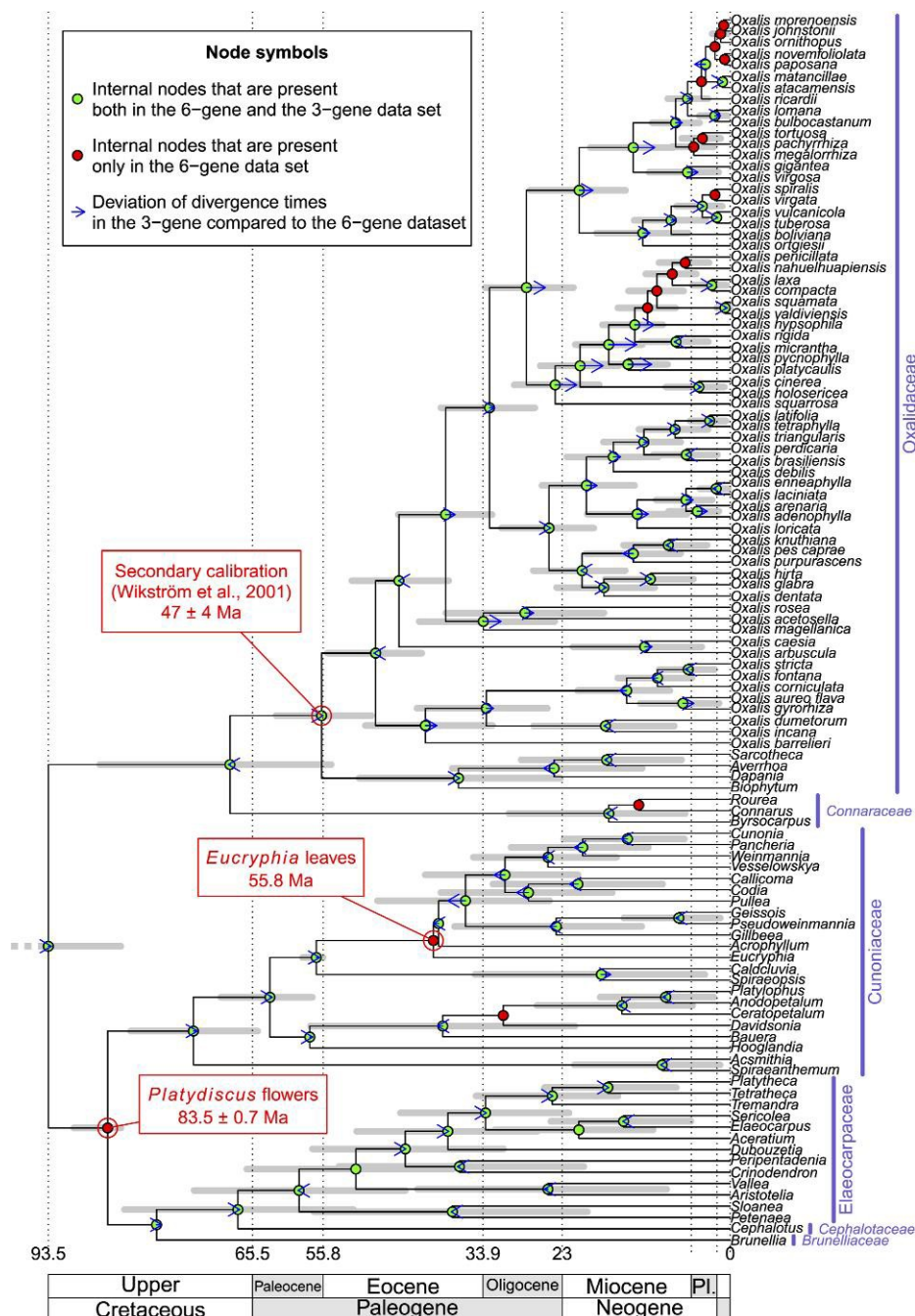


FIGURE 2. Chronogram for the order Oxalidales obtained with a fossil-calibrated relaxed molecular clock model applied to combine nuclear and chloroplast DNA sequences for 112 species (4900 aligned nucleotides of *rbcl*, *trnL-F*, and *ITS*).

distribution, performed well but also predicted occurrences in uninhabitable habitats. The inclusion of fog greatly improved model performance. *Oxalis* species richness is concentrated in Chile's Mediterranean zone at 30–36° S, especially in the Cerro La Campana range (I), the Cerro Cantillana range (II), and the Precordillera (Andean foothills, III; Fig. 5a). A secondary center is the coastal fog belt between Chañaral and Caleta El Cobre (region IV). Lineage richness (Fig. 5b) is highest in the Mediterranean zone and the transition zone to

temperate climate below 600 m, in contrast to species richness, which peaks above timberline (Fig. 5a).

DISCUSSION

The main questions at the outset of this study were how the timing of *Oxalis* diversification in Chile relates to the Andean orogeny and the accompanying climate aridification and whether *Oxalis* adaptation to xeric conditions was temporally congruent with the origin of the

TABLE 1. Divergence time estimates for *Oxalis* and the 7 western South American lineages that are the focus of this study (they are numbered and color coded in Fig. 3)

Clade	Stem group ages (Ma)			Crown group ages (Ma)		
	Mean	95% HPD	PP	Mean	95% HPD	PP
<i>Oxalis</i>	56.02	62.26–49.17	1	48.61	55.24–42.28	1
L1: Trifidus	33.43	42.93–25.09	1	17.09	26.94–7.66	1
L2: Corniculatae p.p.	14.19	20.91–8.28	1	6.48	12.34–1.51	1
L3: Caesiae	45.41	51.87–38.59	0.94	11.72	21.54–3.90	1
L4: Roseae	28.13	38.97–17.31	0.82	—	—	—
L5: Palmatifoliae	19.71	25.81–13.11	1	12.76	19.6–5.91	1
L6: Alpinae	27.97	33.59–21.49	1	23.99	29.6–17.95	1
L7: Camosae + Giganteae	20.70	27.44–14.35	1	13.29	19.19–7.81	1

Note: Mean ages, their 95% highest posterior density interval (HPD), and posterior probability (PP).

TABLE 2. Species richness, habitat, and morphological traits of the 7 *Oxalis* lineages that are the focus of this study and which are numbered and color coded in Figure 3

Lineage	#	% Sampled	Distribution and habitat	Morphology
L1	2(–3)	100(–67)	Disjunction in the Altiplano (>4300 m) and temperate Chile.	Creeping perennial herbs, taproots, leaves tiny with a strong indumentum or large and glabrous.
L2	2	100	Disturbed sites in forests in the Mediterranean–temperate ecotone.	Creeping perennial herbs, taproots and mesic leaves.
L3	2(–4)	100(–50)	Hyperarid sites above the coastal fog layer and in the interior of the Atacama Desert.	Perennial dwarf shrubs, root tubers and microphyllous leaves with thick cuticles.
L4	1	100	Mediterranean and ecotone to temperate, growing in shaded understory.	Fleshy annual herb, fibrous root system and leaves glabrous and delicate.
L5	5	100	Patagonian Andes and cold temperate steppes; one species in Mediterranean Chile	Perennial herbs, fleshy leaf bases forming elongates rhizomes or imbricate bulbs and leaves glabrous to hairy.
L6	24	67	Mediterranean and ecotone to temperate, from sea level to the limit of vegetation and on the eastern slopes in Argentina.	Annual or perennial herbs, alpine representatives cushion-forming, taproots, considerable variation in leaf size and indumentum depending on habitat.
L7	14	94	Major diversity in the coastal fog deserts of Chile and Peru, with extensions into Mediterranean Chile and the Altiplano region.	Perennial (sub-)shrubs, strong taproots or root tubers, stems and leaves fleshy and leaves and flowers borne apically or on short shoots.

Note: Column 2 refers to species number (#) and column 3 to the percentage of species sampled for this study.

respective habitats. We also wanted to know whether Chile's Mediterranean zone, which harbors the greatest species number, is better interpreted as a refuge (i.e., a sink) or a source of lineages. In the following, we first answer these questions and then discuss the effects of the different handling of pseudo-absence data on the performance of the SDMs.

Specifics of *Oxalis* Evolution in Southern South America

The *Oxalis* clade, with about 480 species worldwide, is inferred as having originated in South America. This is implied by non-Neotropical *Oxalis* clades, including the South African clade (marked GCFR in Fig. 3), being nested within South American *Oxalis*. Based on our chronogram (Fig. 2), *Oxalis* had begun radiating well before the first major Andean uplift pulse, and at least 7 lineages occur in Chile that have their closest relatives throughout the genus (Figs. 1, 3, and 4). This is in stark contrast to the *Oxalis* flora of the GCFR, which consists of a single clade of 200 species (Oberlander et al.

2009), the ancestor of which arrived from South America 15–36 Ma (Verboom et al. 2009). Most of the 7 lineages that comprise Chilean *Oxalis* are species-poor, and none show speciation comparable to any of the hyperdiverse Mediterranean climate-adapted South African plant groups (Sauquet et al. 2009; Verboom et al. 2009; Schnitzler et al. 2011).

The uplift of the Andes promoted allopatric speciation in numerous plant clades (Hughes and Eastwood 2006; Antonelli et al. 2009). One might have expected this to be true also for Chilean *Oxalis*. However, although southern South American *Oxalis* lineages entered the alpine habitat 6–8 times (Fig. 4), no high-elevation radiation followed. This is in contrast to the *O. tuberosa* alliance, which comprises about 40 species that diversified in páramo, puna, and cloud forest habitats from northern Argentina to Venezuela (Emshwiller and Doyle 2002). Based on the divergence times obtained here, the *O. tuberosa* radiation took place during and after the second Andean uplift pulse (10–6 Ma), although our sparse species sampling in this

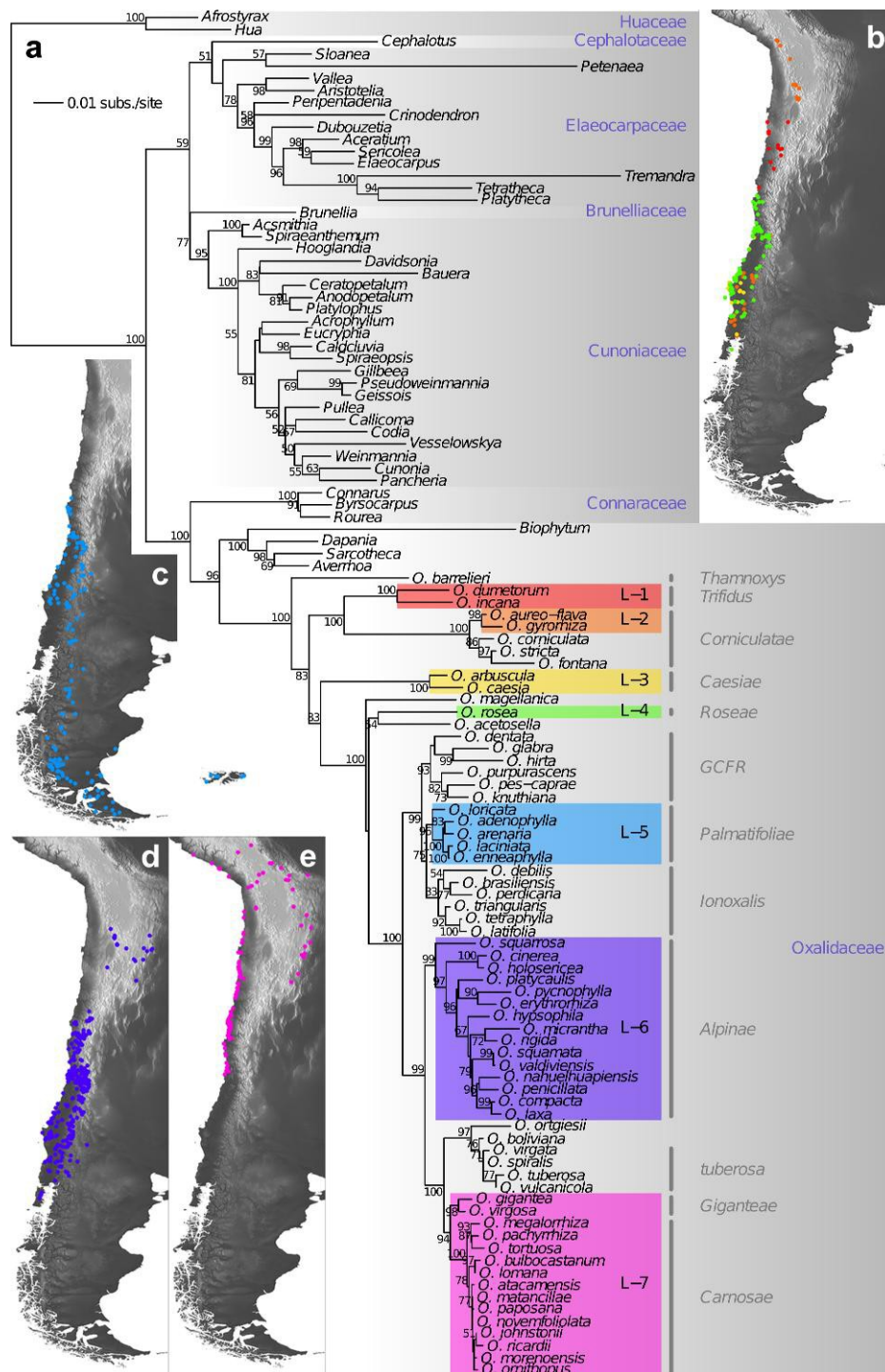


FIGURE 3. Phylogenetic relationships in Oxalidales with the 7 arid western South American lineages of *Oxalis* and their distributional ranges highlighted. (a) ML phylogeny based on the combined chloroplast and nuclear data, with the arid western South American (AWSA) *Oxalis* lineages numbered and color coded. Numbers above branches indicate bootstrap support from 100 replicates. (b-e) Geographic ranges of the 7 AWSA lineages. Dot colors correspond to the color coding in (a). GCFR: Greater Cape floristic region.

non-Chilean clade cautions against overconfidence in this estimate. We also found little evidence of rapid adaptive radiation in the other *Oxalis* clades studied here.

Our reconstructions of habitat preferences revealed no clear temporal coincidence between the availability

of a habitat and its occupation. Instead, the occupation of arid habitats by different groups of *Oxalis* occurred over a long time, probably involving traits, such as tubers, that facilitated shifts to water-stressed habitats. A temporal match between the origin of a habitat type and its occupation exists only in the 2 desert

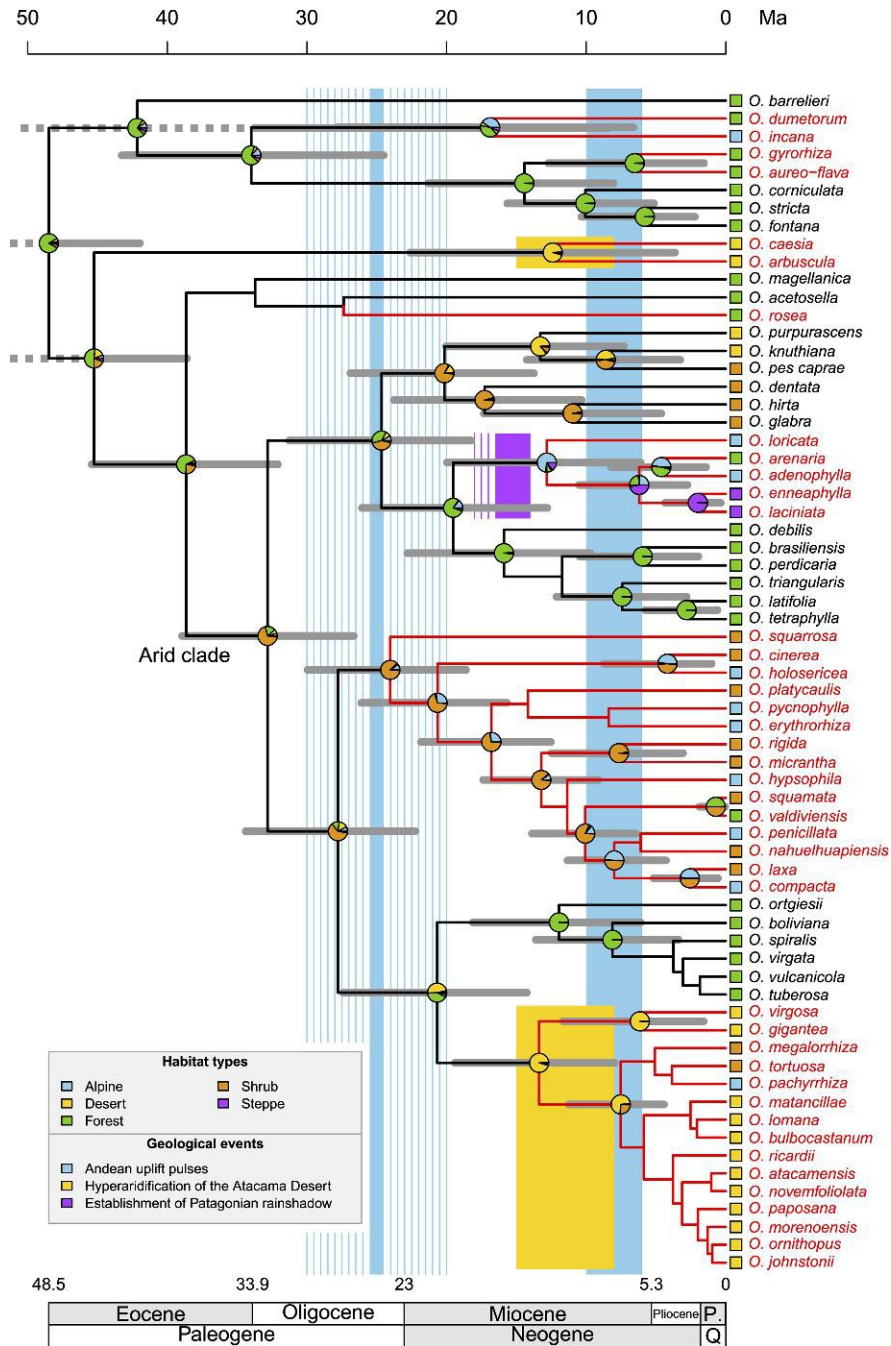


FIGURE 4. Chronogram for *Oxalis* obtained under a Bayesian relaxed clock, with the lineages adapted to arid biomes in the Southern Cone of South America highlighted in red. Bars around node ages indicate the 95% highest posterior density intervals for nodes with a posterior probability >0.9. For each of these nodes, ancestral habitat occupancy obtained under an all-rates-equal ML model is depicted; forest (green), Mediterranean shrub (orange), desert (yellow), alpine (sky blue), and temperate steppe (purple). Vertical bars represent important geological events: Two phases of rapid Andean uplift (sky blue, hatching indicates uncertainty), the hyperaridification of the Atacama Desert (yellow), and establishment of the Patagonian Andean rain shadow (purple, hatching indicates uncertainty).

clades (yellow bands in Fig. 4). Occupation of the Patagonian steppes occurred only in the Pliocene, when this biome would have been available for 8–10 myr. Another habitat shift from forest to Mediterranean shrub land predated the establishment of the shrub biome by 20 myr.

Chile's Mediterranean Climate Zone as a Refuge or Source Area for Lineages of the Atacama and the High Andes

Species distribution modeling identified Chile's Mediterranean zone at 30–36° S, especially the Cerro La Campana range (I), the Cerro Cantillana range (II), and the Precordillera (Andean foothills, III), as the center

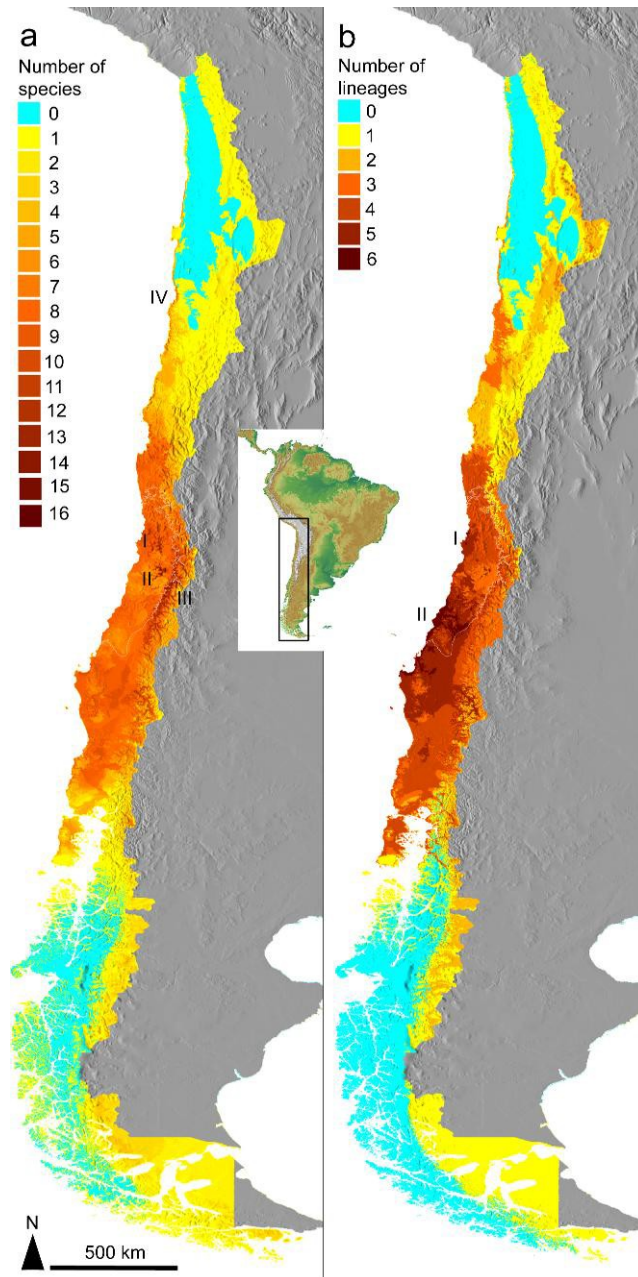


FIGURE 5. Geographic distribution of the diversity of *Oxalis* in Chile as obtained by climatic distribution modeling, with the CAS strategy and fog included (Materials and Methods section). (a) The estimated number of species per landscape unit. (b) The estimated number of lineages per landscape unit, when individual distribution models are pooled per lineage. Pixels are color coded from yellow (low diversity) to dark red (high diversity). A white line indicates the Mediterranean climate zone.

of *Oxalis* species diversity (Fig. 5a). Lineage diversity (Fig. 5b) is highest in the Mediterranean zone and the transition zone to temperate climate below 600 m. However, none of the 7 major lineages (Fig. 3) is restricted to the Mediterranean zone, which thus is a region of species or clade survival, not one of rapid radiation. The mildness of the Quaternary climate cycles in the Mediterranean zone may have facilitated such accumulation of (surviving) species. To the north and south of this zone, pronounced dry–wet cycles during the

Pleistocene and extensive glaciation (Hultón et al. 2002; Maldonado et al. 2005; Kaplan et al. 2008) probably caused more extinction. In a way, the Chilean Mediterranean zone is an *Oxalis* refuge or sink in which species that evolved under different conditions now coexist.

Correcting for Uneven Sampling in Species Distribution Modeling

Unequally distributed collecting can dramatically bias estimates of species ranges and endemism

(e.g., Nelson et al. 1990). Schulman et al. (2007) proposed the use of “CASSs” to account for such collecting biases in the elaboration of SDMs. Their method requires defining a buffer around each collecting point, which is then modified by information from the CAS to obtain a probability surface for the species occurrence (see Schulman et al. 2007 for details). This approach was useful in the Amazon basin, where environmental data are scarce or relatively invariable. In arid western South America, where environmental heterogeneity is extremely high, we relied on climatic and topographic data to develop SDMs but also had to account for sampling bias. The first of the approaches we compared, the overlapping target-group background strategy, performed poorly probably because we had only a fifth of the number of target-group localities recommended for this approach to work well (Phillips et al. 2009). The next 2 approaches, introducing bias files, worked well. We derived 2 such files by either counting accessions per grid cell (NAPG) or by adding a step of nn interpolation to obtain a CAS. Like all raster models, both methods are sensitive to the choice of grid size. A vector-based approach, for example using Thiessen polygons (Schulman et al. 2007), may offer a scale-independent measure of collecting bias and might be worth further exploration.

CONCLUSIONS

Adaptations to aridity, for instance, stunted growth, root tubers, and fleshy leaf bases (Fig. 1), exist in *Oxalis* lineages of widely different phylogenetic relationships and ages and seem to have enabled some populations to take advantage of new ecological opportunities arising with the Late Miocene rain shadow climate. The Mediterranean core zone, just south of the Atacama Desert, offered an ecological refuge for species without xerophytic traits or insufficiently arid-adapted to withstand xeric conditions for long periods. There is no evidence of any rapid adaptive radiation. Spatial models of biodiversity performed best when including fog and collecting activity.

SUPPLEMENTARY MATERIAL

Data related to this study are available in TreeBASE (accession 12267) and the Dryad repository (doi:10.5061/dryad.qp0s0).

FUNDING

The project was supported by the German Research Council (RE 603/5-1).

ACKNOWLEDGMENTS

We thank curators of the herbaria CONC, E, HIP, K, SGO, ULS, and USM for *Oxalis* loans or access to

their collections, M. Weigend, Faculty of Biology, University of Bonn, for *Oxalis* from Peru, and B. Gravendeel, Faculty of Mathematics and Natural Sciences, University of Leiden, for DNA aliquots of outgroup taxa.

REFERENCES

- Ackerly D.D. 2004. Adaptation, niche conservatism, and convergence: comparative studies of leaf evolution in the California chaparral. *Am. Nat.* 163:654–671.
- Alpers C.N., Brimhall G.H. 1988. Middle Miocene climatic change in the Atacama Desert, northern Chile: evidence from supergene mineralization at La Escondida. *Geol. Soc. Am. Bull.* 100:1640–1656.
- Antonelli A., Nylander J.A.A., Persson C., Sanmartín I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106:9749–9754.
- Bahnweg G., Schulze S., Möller E.M., Rosenbrock H., Langebartels C., Sandermann H.J. 1998. DNA isolation from recalcitrant materials such as tree roots, bark, and forest soil for the detection of fungal pathogens by polymerase chain reaction. *Anal. Biochem.* 262:79–82.
- Barnes R.W., Jordan G.J. 2000. Eucryphia (Cunoniaceae) reproductive and leaf macrofossils from Australian Cainozoic sediments. *Aust. Syst. Bot.* 13:373–394.
- Blisniuk P.M., Stern L.A., Chamberlain C.P., Idelman B., Zeitler P.K. 2005. Climatic and ecologic changes during Miocene surface uplift in the Southern Patagonian Andes. *Earth Planet. Sci. Lett.* 230: 125–142.
- Cereceda P., Osses P., Larrain H., Fariás M., Lagos M., Pinto R., Schemenauer R.S. 2002. Advective, orographic and radiation fog in the Tarapacá region, Chile. *Atmos. Res.* 64:261–271.
- Cereceda P., Schemenauer R.S. 1991. The occurrence of fog in Chile. *J. Appl. Meteorol.* 30:1097–1105.
- Clarke J.D.A. 2006. Antiquity of aridity in the Chilean Atacama Desert. *Geomorph.* 73:101–114.
- Crisp M.D., Arroyo M.T.K., Cook L.G., Gandolfo M.A., Jordan G.J., McGlone M.S., Weston P.H., Westoby M., Wilf P., Linder H.P. 2009. Phylogenetic biome conservatism on a global scale. *Nature.* 458:754–756.
- Dillon M.O., Tu T., Lei X., Quipuscoa Silvestre V., Wen J. 2009. Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian Deserts along the western coast of South America. *J. Syst. Evol.* 47:457–476.
- Donoghue M. 2009. A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* 105:11549–11555.
- Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:240.
- Emshwiller E., Doyle J.J. 2002. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*: Oxalidaceae). 2. Chloroplast-expressed glutamine synthase data. *Am. J. Bot.* 89:1042–1056.
- Fuentes E.R., Montenegro G., Rundel P.W., Arroyo M.T.K., Ginocchio R., Jaksic F.M. 1995. Function approaches to biodiversity in the Mediterranean-type ecosystems of central Chile. In: Davis G.W., Richardson D.M., editors. *Mediterranean-type ecosystems: the function of biodiversity*. vol. 109. Berlin (Germany): Springer. p. 185–232.
- GRASS Development Team 2009. Geographic Resources Analysis Support System (GRASS GIS) Software. Open Source Geospatial Foundation. Available from: <http://grass.osgeo.org/impressum.html>.
- Hamilton M.B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol.* 8:521–523.
- Hartley A.J., Chong G., Houston J., Mather A.E. 2005. 150 million years of climatic stability: evidence from the Atacama Desert, northern Chile. *J. Geol. Soc.* 162:421–425.
- Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965–1978.

- Houston J., Hartley A.J. 2003. The central Andean west-slope rain-shadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert. *Int. J. Climatol.* 23:1453–1464.
- Hughes C., Eastwood R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl. Acad. Sci. U.S.A.* 103:10334–10339.
- Hulton N.R.J., Purves R.S., McCulloch R.D., Sugden D.E., Bentley M.J. 2002. The Last Glacial Maximum and deglaciation in southern South America. *Quat. Sci. Rev.* 21:233–241.
- Jiménez-Valverde A., Lobo J.M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecol.* 31:361–369.
- Kaplan M.R., Fogwill C.J., Sugden D.E., Hulton N.R.J., Kubik P.W., Freeman S.P.H.T. 2008. Southern Patagonian glacial chronology for the Last Glacial period and implications for Southern Ocean climate. *Quat. Sci. Rev.* 27:284–294.
- Larriain H., Velásquez F., Cereceda P., Espejo R., Pinto R., Osses P., Schemenauer R.S. 2002. Fog measurements at the site “Falda Verde” north of Chañaral compared with other fog stations of Chile. *Atmos. Res.* 64:273–284.
- Lourteig A. 2000. *Oxalis* L. subgéneros *Monoxalis* (Small) Lourt., *Oxalis* y *Trifidus* Lourt. *Bradea*. 7:201–628.
- Löytynoja A., Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U.S.A.* 102:10557–10562.
- Luebert F., Pliscoff P. 2006. Sinopsis bioclimática y vegetacional de Chile. Santiago de Chile, Chile: Editorial Universitaria.
- Luebert F., Wen J. 2008. Phylogenetic analysis and evolutionary diversification of *Heliotropium* sect. *Cochranea* (Heliotropiaceae) in the Atacama Desert. *Syst. Bot.* 33:390–402.
- Maldonado A., Betancourt J.L., Latorre C., Villagrán C. 2005. Pollen analyses from a 50000-yr rodent midden series in the southern Atacama Desert (25° 30'S). *J. Quat. Sci.* 20:493–507.
- Nelson B.G., Ferreira C.A.C., da Silva M.F., Kawasaki M.L. 1990. Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature*. 345:714–716.
- Oberlander K.C., Emshwiller E., Bellstedt D.U., Dreyer L.L. 2009. A model of bulb evolution in the eudicot genus *Oxalis* (Oxalidaceae). *Mol. Phylogenet. Evol.* 51:54–63.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B Biol. Sci.* 255:37–45.
- Pagel M., Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* 167:808–825.
- Pagel M., Meade A., Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53:673–684.
- Peterson A.T., Soberon J., Sanchez-Cordero V. 1999. Conservatism of ecological niches in evolutionary time. *Science*. 285:1265–1267.
- Phillips S.J., Anderson R.P., Schapire R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190:231–259.
- Phillips S.J., Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*. 31:161–175.
- Phillips S.J., Dudík M., Elith J., Graham C.H., Lehmann A., Leathwick J., Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19:181–197.
- Prinzing A., Durka W., Klotz S., Brandl R. 2001. The niche of higher plants: Evidence for phylogenetic conservatism. *Proc. R. Soc. Lond. B Biol. Sci.* 268:2383–2389.
- R Development Core Team 2008. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rambaut A., Drummond A.J. 2007. Tracer v1.4. Available from: <http://beast.bio.ed.ac.uk/Tracer>.
- Rech J.A., Currie B.S., Shullenberger E.D., Dunagan S.P., Jordan T.E., Blanco N., Tomlinson A.J., Rowe H.D., Houston J. 2010. Evidence for the development of the Andean rain shadow from a Neogene isotopic record in the Atacama Desert, Chile. *Earth Planet. Sci. Lett.* 292:371–382.
- Rundel P.W., Dillon M., Palma B., Mooney H., Gulmon S., Ehleringer J. 1991. The phytogeography and ecology of the coastal Atacama and Peruvian deserts. *Aliso*. 13:1–49.
- Sakov P. 2010. nn—natural neighbours interpolation library, version 1.82. Available from: <http://code.google.com/p/nn-c/>.
- Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19:101–109.
- Sang T., Crawford D.J., Stuessy T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* 84:1120–1136.
- Sauquet H.I., Weston P.H., Anderson C.L., Barker N.P., Cantrill D.J., Mast A.R., Savolainen V. 2009. Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 106:221–225.
- Schmitz J., Barraclough T.G., Boatwright J.S., Goldblatt P., Manning J.C., Powell M.P., Rebelo T., Savolainen V. 2011. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Syst. Biol.* 60:343–357.
- Schönenberger J., Friis E.M., Matthews M.L., Endress P.K. 2001. Cunoniaceae in the Cretaceous of Europe: evidence from fossil flowers. *Ann. Bot.* 88:423–437.
- Schulman L., Toivonen T., Ruokolainen K. 2007. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *J. Biogeogr.* 34:1388–1399.
- Sieczka M. 2008. r.surf.nnbathy—interpolate surface from rasterized input using Pavel Sakov's nn, a natural neighbor interpolation library. Available from: <http://www.sieczka.org/prog/grass/r.surf.nnbathy.html>.
- Stamatakis A. 2006. RAXML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–2690.
- Stamatakis A., Hoover P., Rougemont J. 2008. A rapid bootstrap algorithm for the RAXML web-servers. *Syst. Biol.* 57:758–771.
- Taberlet P., Gielly L., Pautou G., Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17:1105–1109.
- Talavera G., Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56:564–577.
- USGS 1996. GTOPO30: a global 30 arc-second elevation data set. Available from: http://eros.usgs.gov/-/Find_Data/Products_and_Data_Available/gtopo30.info.
- Verboom G.A., Archibald J.K., Bakker F.T., Bellstedt D.U., Conrad F., Dreyer L.L., Forest F., Galley C., Goldblatt P., Henning J.F., Mummenhoff K., Linder H.P., Muasya A.M., Oberlander K.C., Savolainen V., Snijman D.A., Niet T. van der, Nowell T.L. 2009. Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Mol. Phyl. Evol.* 51:44–53.
- Webb C.O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am. Nat.* 156:145–155.
- Wiens J.J., Graham C.H. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Ann. Rev. Ecol. Evol. Syst.* 36:519–539.
- Wikström N., Savolainen V., Chase M.W. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B Biol. Sci.* 268:2211–2220.